

RESEARCH ARTICLE

Analysis of bacterial genomes from an evolution experiment with horizontal gene transfer shows that recombination can sometimes overwhelm selection

Rohan Maddamsetti^{1,2,3*}, Richard E. Lenski^{1,2,4}

1 Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, MI, United States of America, **2** BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI, United States of America, **3** Department of Systems Biology, Harvard Medical School, Boston, MA, United States of America, **4** Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, United States of America

* rohan@hms.harvard.edu.



 OPEN ACCESS

Citation: Maddamsetti R, Lenski RE (2018) Analysis of bacterial genomes from an evolution experiment with horizontal gene transfer shows that recombination can sometimes overwhelm selection. *PLoS Genet* 14(1): e1007199. <https://doi.org/10.1371/journal.pgen.1007199>

Editor: Ivan Matic, Université Paris Descartes, INSERM U1001, FRANCE

Received: October 26, 2017

Accepted: January 15, 2018

Published: January 31, 2018

Copyright: © 2018 Maddamsetti, Lenski. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Genome sequence data have been deposited in the NCBI database under the study accession SRP114715, with the respective BioSample accessions SAMN07440982 (donors, recipients, and clones REL4397 and REL4398 from population Ara-3), SAMN07440983 (recombinant clones), and SAMN07440984 (mixed populations sampled at the start and end of the STLE continuation experiment), under BioProject PRJNA396011. All data and analysis code for this project, excluding the genome sequencing read files, are available at the Dryad digital repository

Abstract

Few experimental studies have examined the role that sexual recombination plays in bacterial evolution, including the effects of horizontal gene transfer on genome structure. To address this limitation, we analyzed genomes from an experiment in which *Escherichia coli* K-12 Hfr (high frequency recombination) donors were periodically introduced into 12 evolving populations of *E. coli* B and allowed to conjugate repeatedly over the course of 1000 generations. Previous analyses of the evolved strains from this experiment showed that recombination did not accelerate adaptation, despite increasing genetic variation relative to asexual controls. However, the resolution in that previous work was limited to only a few genetic markers. We sought to clarify and understand these puzzling results by sequencing complete genomes from each population. The effects of recombination were highly variable: one lineage was mostly derived from the donors, while another acquired almost no donor DNA. In most lineages, some regions showed repeated introgression and others almost none. Regions with high introgression tended to be near the donors' origin of transfer sites. To determine whether introgressed alleles imposed a genetic load, we extended the experiment for 200 generations without recombination and sequenced whole-population samples. Beneficial alleles in the recipient populations were occasionally driven extinct by maladaptive donor-derived alleles. On balance, our analyses indicate that the plasmid-mediated recombination was sufficiently frequent to drive donor alleles to fixation without providing much, if any, selective advantage.

Author summary

Bacteria often transfer genes encoding antibiotic resistance as well as other important traits, but the extent of intergenomic recombination—in effect, sex—is highly variable

[doi: 10.5061/dryad.3bv80]. The code for the analysis is also available on GitHub (www.github.com/rohanmaddamsetti/STLE-analysis).

Funding: This work was supported by a grant from the National Science Foundation (DEB-1451740 to REL), the USDA National Institute of Food and Agriculture (MICL02253 to REL), the BEACON Center for the Study of Evolution in Action (NSF Cooperative Agreement DBI-0939454), a National Defense Science and Engineering Graduate Fellowship (to RM), and a dissertation completion fellowship from the MSU Graduate School (to RM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

across bacterial species. Why? A better understanding of how and why bacteria exchange genes would help people combat the spread of infectious disease as well as shed light on the evolutionary origins of sex. Here, we sequenced genomes from an evolution experiment with *Escherichia coli* in which recombination was extensive but, unexpectedly, did not speed up the rate of adaptation. In this experiment, the effective rate of chromosomal recombination was much higher than previously inferred for natural *E. coli* populations. In fact, the rate was so high that introduced genes sometimes drove established beneficial alleles to extinction in the experimental populations. In effect, genes that were physically linked to the genes causing recombination had a strong transmission advantage, whether or not they provided any selective advantage to the recipient cells.

Introduction

An open question in microbial evolution is why some bacterial taxa seem to have extensive intergenomic recombination [1] while others seem to have little [2]. Some symbiotic bacteria appear to be entirely clonal owing to their tight associations with hosts that preclude contact with other lineages [3], but otherwise the reasons for differences in recombination rates across bacterial taxa are unclear. Intergenomic recombination can break the linkage between particular beneficial or deleterious mutations and the rest of the genome. Under conditions of high recombination and strong selection, individual genes, rather than entire genomes, can thus go to fixation. When recombination is infrequent or absent but selection is strong, highly beneficial mutations can drive large genomic regions or even whole genomes to fixation. Recent work shows that both gene-specific and genome-wide selective sweeps occur in microbial communities [4]. In contrast to meiotic recombination, however, bacterial recombination replaces a recipient allele with a donor allele, rather than swapping homologous regions between chromosomes. Thus, recombination in bacteria does not necessarily break up linkage disequilibrium across distant sites on the recipient chromosome; instead, it may preserve a “clonal frame” over most of the chromosome, interrupted by stretches of DNA introduced by horizontal gene transfer [5].

Horizontally transmitted viruses and conjugative plasmids mediate recombination in many species of bacteria [6]. In this context, recombination is a special kind of evolutionary process. Like mutation rates, intergenomic recombination rates can evolve because many gene products involved in DNA replication and repair affect the rates of mutation and recombination [7, 8]. But unlike mutation rates, recombination rates in bacteria are also subject to coevolution owing to the association with plasmids and viruses [9]. Intergenomic recombination qualitatively changes evolutionary dynamics [10, 11] and can speed up adaptive evolution compared to asexual controls under some circumstances [12, 13]. However, recombination—especially as it occurs in bacteria—may have originated as a byproduct of the spread of selfish elements rather than as a means of increasing the efficiency of natural selection in the genome as a whole [14].

The conjugative element relevant to our study is called the F plasmid. It encodes proteins that form a pilus between an F⁺ donor cell and an F⁻ recipient cell. F also encodes proteins that transfer the plasmid (and sometimes host DNA) to the recipient after cell-cell contact has been made. Strains that have the F plasmid integrated into their chromosome are called Hfr (high frequency of recombination) strains. The F plasmid can sometimes spontaneously excise from the Hfr chromosome; if a piece of the host chromosome is excised along with the F plasmid itself, then those cells are called F' to indicate that the plasmid also contains bacterial DNA.

The location and orientation of the *oriT* site in the chromosome of an Hfr strain determines the order and timing of the transfer of donor DNA into the recipient cell. Sequences close to *oriT* and in the proper orientation are transferred first, whereas sequences far from *oriT* are transferred only later if at all [15, 16].

In this paper, we revisit an evolution experiment in which sexual recombination did not appear to increase the efficiency of natural selection, nor did it speed up adaptation, contrary to some hypotheses about the evolutionary advantages associated with sex. Souza, Turner, and Lenski [17] conducted an experiment in which they periodically introduced an equal mixture of 4 Hfr strains of *Escherichia coli* K-12 that could donate genetic material but not grow (owing to mutations that caused nutritional deficiencies) into 12 recipient populations of *E. coli* B, each founded by single clones that had previously evolved in and adapted to a glucose-limited minimal medium for 7000 generations. Each Hfr strain had an F plasmid inserted at a different position and orientation in its chromosome, and the same mix of 4 Hfr strains was used for all 12 recipient populations throughout the experiment. We refer to the recombination treatment in that study as the Souza-Turner-Lenski experiment (STLE) and to the experiment that generated the asexual recipients as the long-term evolution experiment (LTEE). The stated goal of the STLE was to test the hypothesis that recombination would increase the rate of adaptation by increasing the genetic variation available to natural selection [17]. In the absence of complex selection dynamics (such as frequency-dependent selection), the expected rate of adaptation of a population is proportional to the genetic variance in fitness in a population [18]. The outcome of the STLE was unexpected in that the recombination treatment increased genetic variation (as determined by tracking ~10 genetic markers available to the authors at that time), but it had no significant effect on the rate of adaptation compared to control populations that evolved without recombination.

Souza *et al.* (1997) proposed three hypotheses that could explain their puzzling results. These hypotheses are not mutually exclusive. According to one, the recombination treatment was so effective that it acted like a strong mutational force, replacing many neutral alleles (or even overwhelming selection, if donor alleles were maladaptive) through the sheer flux of donor DNA into the recipient genomes. According to another, some donor alleles hitchhiked to high frequency with beneficial 'driver' mutations that arose in the recipient populations. Lastly, the interactions between donor and recipient genes, and the ecological context in which those interactions occurred, might have somehow generated strong frequency-dependent selection, such that the assays used to measure fitness gains were undermined. In fact, one recombinant population of the STLE actually appeared to decline substantially in fitness, an effect that was shown to reflect an evolved frequency-dependent interaction [19]. However, this effect was tested in only one population, and no evidence bearing on the other two hypotheses was available. In this study, we characterize the genomic evolution that occurred during the STLE, with the aim of resolving the puzzling results of the STLE. On balance, we find substantial evidence that the recombination treatment acted like a greatly increased mutational force.

Results

Structure of recombinant genomes

Fig 1 summarizes the rich and complex information on genomic changes that occurred during the 1000-generation STLE. Two clones were sampled at random from each of the 12 populations in the recombination treatment, and their genomes were sequenced and analyzed along with those of the donors and LTEE-derived recipients. The odd-numbered clones from each population are shown in Fig 1; the even-numbered clones are shown in S2 Fig. The genomic sites marked in blue show mutations that distinguish the 12 clones that were used as recipients

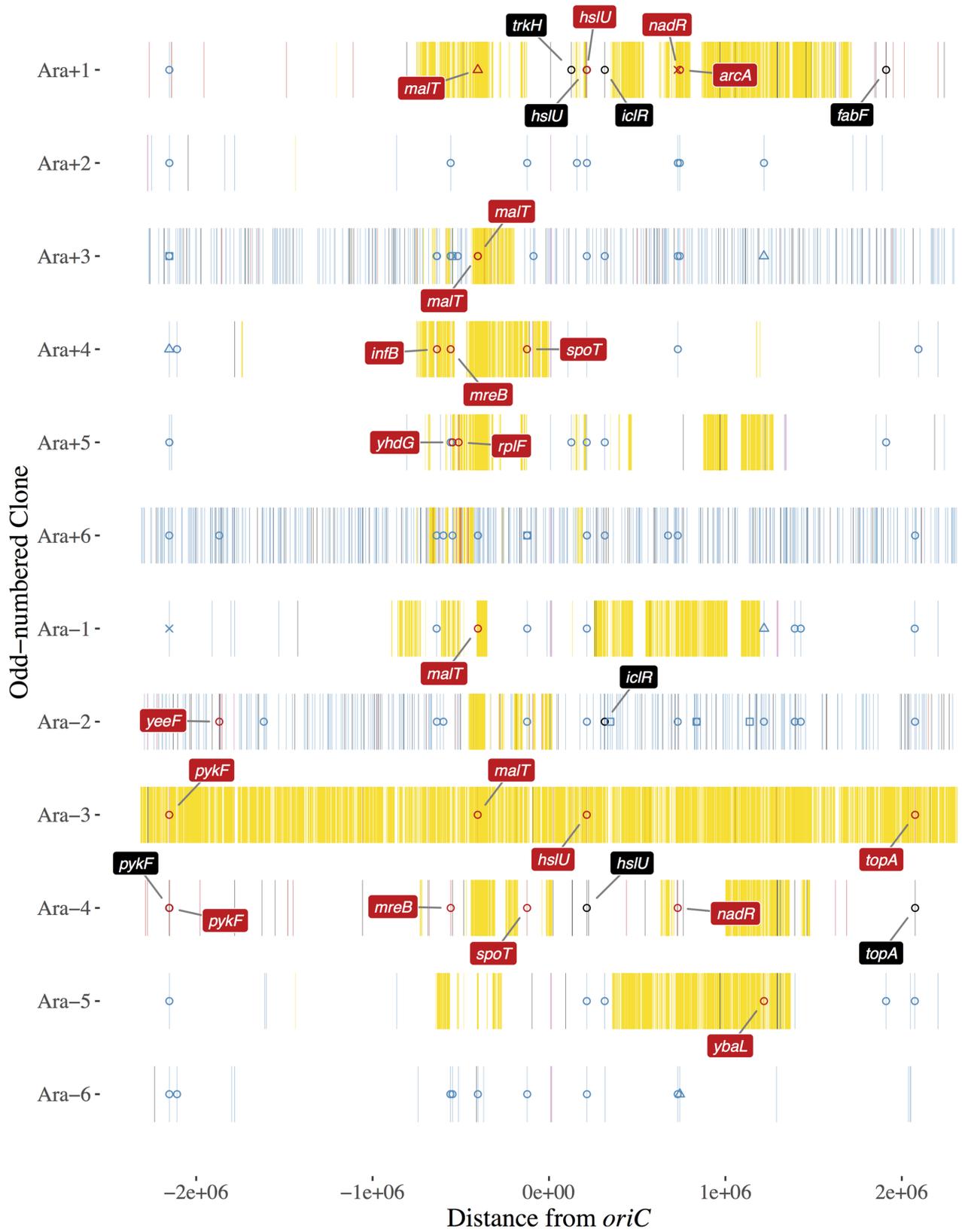


Fig 1. Genome structure of odd-numbered clones from recombinant populations after 1000 generations of the STLE. The REL606 genomic coordinates are shown on the x-axis, centered on the *oriC* origin of replication, and the source populations are shown on the y-axis. Genetic markers are shown as vertical lines, with the color indicating the origin of each marker. Markers specific to K-12 donors are yellow; markers specific to recipient clones are blue; markers in deleted regions are light purple; new mutations that arose during the STLE are black; and LTEE-derived mutations that were replaced by donor DNA during the STLE are red. In addition, symbols indicate mutations in genes under positive selection in the LTEE (Table 1). Open circles indicate nonsynonymous point mutations; open squares are synonymous mutations; open triangles are indels; and x-marks are IS-element insertions. Replaced and new mutations in the genes in Table 1 are labeled by their gene names.

<https://doi.org/10.1371/journal.pgen.1007199.g001>

in the STLE from the ancestor of the LTEE, and which were still present in the genome of a recombinant clone from the STLE. These mutations thus arose during the 7000 generations of the LTEE that preceded the start of the STLE, and they persisted for the 1000 generations of the STLE. The clones from populations Ara+3, Ara+6, and Ara-2 have many more such mutations than the clones from the other nine STLE populations, because the recipient clones used to start these populations came from populations that had evolved hypermutable phenotypes during the early generations of the LTEE [20, 21]. The genomic sites marked in yellow are mutations shared by the K-12 donor strains and the sequenced recombinant clone, but which were not present in the recipient clone used to start that population. These K-12 alleles were thus introduced by intergenomic recombination during the STLE. For comparison, Fig 2 shows the location and direction of the Hfr origins of transfer of the four K-12 donor strains, and the location of their auxotrophy mutations, which together explain much of the overall pattern of K-12 introgression.

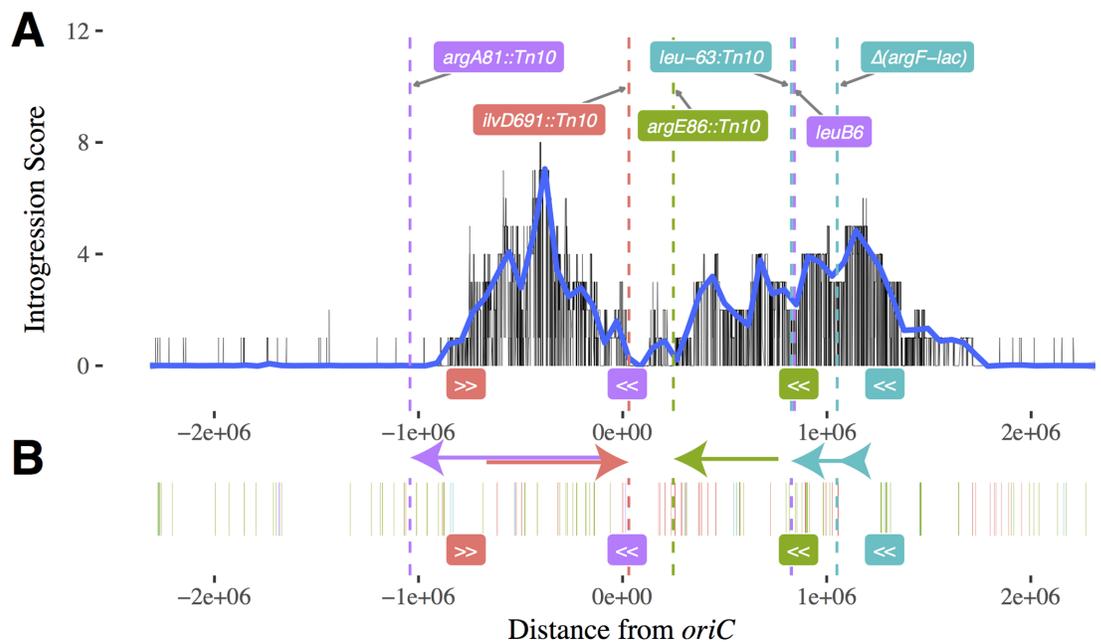


Fig 2. Impact of donors' transfer origins and auxotrophic mutations on introgression. (A) The number of parallel introgressions of K-12 genetic markers summed over the odd-numbered STLE clones (omitting the Ara-3 clone, which is almost completely derived from K-12 donor DNA). A natural cubic spline with 100 degrees of freedom, in blue, was fit to these data. The locations of auxotrophic mutations in the donor genomes are shown as dashed vertical lines, and the location and orientation of the *oriT* transfer origin sites are labeled below the x-axis. The markers and transfer origins for REL288 are shown in red, for REL291 in green, for REL296 in teal, and for REL298 in purple. The colored arrows indicate the genome regions most likely to be transferred by each Hfr. (B) Auxotrophic mutations do not cause a complete barrier to introgression. The location of donor-specific markers found in the odd-numbered recombinant clones (including Ara-3) are labeled with the same colors as in panel (A). Note that almost all of the introgressions near an auxotrophic mutation in one donor strain came from a different donor that did not carry that mutation.

<https://doi.org/10.1371/journal.pgen.1007199.g002>

These differentially marked sites reveal several features. First, the majority of sites in most genomes derive from the recipients, not from the donor strains. In fact, an individual clone from population Ara-6 (Fig 1) appears to lack any DNA regions that derive from the donors, and both clones from population Ara+2 only have one very short donor segment (~1 kbp in length, barely visible near -1.3 Mbp in Fig 1). Second, there is one striking exception to the above pattern: the genomes of both clones from population Ara-3 are largely comprised of donor-derived DNA (Fig 1). Although similar to one another at a coarse-grained level, these clones still differ by more than 400 K-12 alleles. We also sequenced the genomes of two other clones (REL4397 and REL4398) from this same population that were used in a previous study of frequency-dependent selection that demonstrated ecological differences between these clones [19]. They, too, are predominantly K-12 in origin, but with many small regions that descend from the LTEE recipient clone (S3 Fig). Third, in most STLE populations, the pattern of introgression of donor DNA is very similar in the two recombinant clones we sequenced. However, there are differences in several cases including Ara-6 where, as noted above, one clone appears to lack any donor DNA; Ara+3, where only one clone has any sizable regions of donor DNA; and Ara+4 and Ara-5, where the pairs of recombinant clones share some regions of donor DNA but not others (Fig 1 and S2 Fig).

Fourth, there is an almost complete absence of donor DNA in all of the recombinant populations (except Ara-3, which has mostly donor DNA) in a span of over 2 Mbp (between about +1.8 to -1.0 Mbp on the circular genomic map). Fig 2A shows this point clearly as the sum of the number of introgressions of donor DNA into the odd-numbered clone from each STLE population, excluding the aberrant Ara-3. This region does not reflect a paucity of mutations that could distinguish the donor-derived and recipient-derived DNA, which are as abundant there as elsewhere in the genome (S1 Fig). Fig 2A also shows that donor DNA appears to be concentrated in two regions of the recombinant genomes. One region is centered at about -0.5 Mbp on the map, falling off more or less symmetrically on either side; the other donor-rich region peaks at ~1.1 Mbp to ~1.2 Mbp and extends farther to the left.

In addition to yellow and blue marks that indicate donor- and LTEE-derived alleles, respectively, Fig 1 also has some black, light-purple, and red marks. The black marks indicate mutations that do not exist in either the donor pool or the recipient clone that was used to start a given population. These marks therefore indicate new mutations that arose during the 1000 generations of the STLE. Not surprisingly, there are many more new mutations in the three populations founded by the hypermutable recipient clones. The light-purple marks indicate regions that were deleted in the recombinant genomes. The red marks indicate mutations present in the LTEE-derived recipient clone but absent from the STLE-derived recombinant clone. These marks imply that the mutations that evolved at those sites during the LTEE were effectively “erased”, being replaced by donor alleles during the STLE. Most of these replaced mutations are surrounded by yellow, showing that they resulted from recombination. However, some of them are not surrounded by yellow; for example, the even-numbered Ara+3 clone has several red marks but no yellow marks (S2 Fig). Nonetheless, these isolated red marks also probably reflect recombination, not mutation, given the experimental treatment with Hfr strains. Note that if there were no donor-specific markers closely adjacent to a small region of introgressed DNA, then we could not unambiguously assign a segment as donor-derived.

Probable beneficial mutations

Fig 1 also has symbols and, in some cases, labels showing the names of certain genes that underwent changes in a recombinant clone. (S2 Fig shows the same information for the other

recombinant clone from each population.) The mutations are marked by symbols that are colored in a similar manner to the lines: blue symbols indicate mutations that were present in the LTEE-derived recipient and retained by the recombinant clone; black symbols are new mutations in the recombinant that were not present in either the recipient or the donor; and red symbols are mutations that were present in the recipient but absent from the recombinant, because they were replaced by donor DNA.

A total of 32 different genes are labeled in one or more recombinant populations. They have been called out because they were probable targets of positive selection under the conditions of the LTEE. Table 1 provides additional information on each of them. These genes were

Table 1. Thirty-two genes under positive selection in the LTEE that harbored mutations in at least one of the recipient clones used to start the STLE.

Index	Gene name	Start position	Coding length	G score
1	<i>yabB</i>	92438	459	14.48
2	<i>fisI</i>	94217	1767	16.07
3	<i>queA</i>	393434	1071	11.09
4	<i>ybaL</i>	473629	1677	9.30
5	<i>mrdA</i>	648900	1902	48.33
6	<i>nagC</i>	682469	1221	10.57
7	<i>fabF</i>	1166508	1242	10.50
8	<i>topA</i>	1329420	2598	52.38
9	<i>sapF</i>	1350197	807	12.23
10	<i>pykF</i>	1732965	1413	180.42
11	<i>infC</i>	1777363	435	14.70
12	<i>yeeF</i>	2016147	1359	17.65
13	<i>atoC</i>	2268566	1382	17.55
14	<i>infB</i>	3248576	2673	43.57
15	<i>nusA</i>	3251273	1488	9.78
16	<i>arcB</i>	3285924	2337	14.39
17	<i>mreC</i>	3326766	1103	27.50
18	<i>mreB</i>	3327935	1044	46.77
19	<i>yhdG</i>	3338171	966	11.51
20	<i>rpsD</i>	3368960	621	13.28
21	<i>rplF</i>	3373512	534	33.30
22	<i>malT</i>	3481685	2706	51.73
23	<i>spoT</i>	3760757	2109	113.18
24	<i>yicL</i>	3798180	924	11.69
25	<i>trkH</i>	4011518	1452	25.30
26	<i>yihP</i>	4044063	1407	10.00
27	<i>hslU</i>	4099899	1332	93.73
28	<i>iclR</i>	4201735	825	127.57
29	<i>plsB</i>	4232975	2424	7.83
30	<i>hsdM</i>	4559434	1586	9.52
31	<i>nadR</i>	4615529	1233	106.19
32	<i>arcA</i>	4627750	717	40.91

Note: Start position is relative to the reference genome of REL606, and coding length is the number of nucleotides. The G score indicates the strength of evidence for positive selection based on parallel evolution (Tenaillon *et al.* 2016).

<https://doi.org/10.1371/journal.pgen.1007199.t001>

previously identified (along with others that did not undergo any changes in the recombinant clones in our study) by sequencing a total of 264 genomes from the 12 LTEE populations at various times through 50,000 generations, and finding that they had accumulated an unexpectedly large number of independent nonsynonymous mutations in lineages that had not become hypermutable [21]. The G scores shown in Table 1 indicate the strength of the evidence for excessive parallelism in a gene, relative to the length of its coding sequence. For some of them, genetic manipulations and competition assays have directly confirmed that mutations indeed improve fitness under the conditions of the LTEE [22, 23].

The STLE's 1000-generation duration is short relative to the LTEE, and so we might not expect to see many new beneficial mutations rising to high frequency in these genes. Furthermore, the starting clones of the STLE had already evolved in and adapted to LTEE conditions for 7000 generations. However, we see many examples including four in STLE population Ara +1 in the *fabF*, *trkH*, *hslU*, and *iclR* genes and three in population Ara-4 in the *topA*, *pykF*, and *hslU* genes (Fig 1). These are also the two non-mutator populations that underwent the most replacements of LTEE-evolved mutations by donor alleles (red hash marks), although we do not know whether this relation is coincidental or meaningful. In the next section, we consider the fate of the presumptively beneficial mutations that were present in the LTEE-derived recipients at the start of the STLE.

Possible sources of variation in introgression across genomic regions, and the fate of previously evolved beneficial mutations

What causes the variation across the genome in the extent of introgression of the Hfr donors' DNA into the recombinant populations? There are several distinct hypotheses that rely either on differences in the propensity for genomic regions to be transferred by the donors or on the fitness effects of integrating different regions into the recipient's chromosome. These hypotheses are not mutually exclusive, and so two or more of them may contribute to the observed patterns of introgression (Fig 1 and Fig 2). Hypothesis 1: Some regions of donor DNA were transferred more often than other regions, leading to overrepresentation of the former regions in the recombinant genomes. Hypothesis 2: Some regions of donor DNA contained alleles that were beneficial to the recipient, leading to overrepresentation of those regions in the recombinant genomes. Hypothesis 3: Some regions of donor DNA contained alleles that were deleterious to the recipient, leading to underrepresentation of those regions in the recombinant genomes. This hypothesis can be subdivided into two variant hypotheses. According to Hypothesis 3A, the donor alleles were maladaptive regardless of the beneficial mutations that arose during the LTEE. According to Hypothesis 3B, the donor alleles were maladaptive specifically because the recipient genomes had acquired beneficial mutations in those regions during the 7000 generations of the LTEE that preceded the STLE.

In addition to the hypotheses above, sequence divergence and mismatch repair are known to reduce recombination efficiency [24]. However, *E. coli* K-12 (the strain that gave rise to the Hfr donors) and B (the strain from which the recipients derive) are fairly closely related, at least as far as *E. coli* strains go. These two source strains were independently isolated from nature many years ago [25]. They differ by ~33,000 mutations (using REL606 as a reference genome), which equals ~8 mutations per kilobase (S1 Fig). Still, about half of their shared genes encode proteins that have identical amino-acid sequences [26]. On the other hand, several hundred genes are present in only one or the other strain, including so-called "genomic islands" that are thought to have been acquired by horizontal gene transfer in the phylogenetic networks leading to one or the other strain [26]. In addition to these more or less ancient differences, the four K-12 donors were deliberately modified by transposon mutagenesis to make

them auxotrophic for different amino acids and by introducing the F plasmid at different locations into their chromosomes to make them Hfr strains; and, as described above, the B-derived recipients accumulated beneficial mutations during the LTEE. Overall, regions containing introgressed K-12 alleles in the STLE-evolved genomes were no more divergent than regions without K-12 alleles (Kruskal-Wallis test, $P = 0.9235$). However, we saw a positive correlation between sequence divergence and the location of recombination breakpoints (Kruskal-Wallis test, $P = 0.0199$) (S5 Fig). Therefore, it appears that sequence divergence had a weak effect on the introgression of K-12 alleles into the STLE populations, which mostly influenced the fine-scale mosaic structure of recombinant regions.

Hypothesis 2 was, in essence, the original motivation for the STLE, with Souza *et al.* (1997) suggesting that intergenomic recombination with the K-12 donors might increase the rate of adaptation (relative to control populations that evolved asexually) by providing an additional source of genetic variation to the LTEE-derived populations. We lack *a priori* information about what sites in the K-12 donor genomes could provide beneficial alleles to the recombinant populations. However, if such sites exist, then we would expect them to be in those regions where the introgression scores are high (Fig 2). On the other hand, Hypothesis 2 seems unlikely, because Souza *et al.* (1997) found that fitness gains were not greater in the recombinant populations than in the control populations, which implies that intergenomic recombination did not increase the supply of beneficial alleles.

Fig 2 shows the inferred location and direction of the Hfr origins of transfer of the four K-12 donor strains as well as the location of their auxotrophy mutations, which bear on Hypotheses 1 and 3A, respectively. With respect to Hypothesis 1, Hfr strains transfer their DNA in a unidirectional manner, and the probability that donor genes are transferred to recipients is expected to decline from about 10^{-2} or 10^{-3} at a distance of 100 kbps from the transfer origin to $\sim 10^{-6}$ at a distance of 2.3 Mbps from the transfer origin [16]. In the STLE, the cultures in which the donors and recipients were mixed were placed in a non-shaking incubator at 37°C for one hour [17]; that duration would, in principle, allow the transfer of $\sim 60\%$ of the entire chromosome if a conjugative mating began immediately after the strains were mixed. However, not all matings would begin immediately, shaking may disrupt conjugation, and the efficiency of DNA transfer declines with distance from the transfer origin. The peak in the introgression scores between about -1 and 0 Mbp fits very well with the locations and directions of the *oriT* transfer origin sites for two of the Hfr donors: REL288 and REL298 have *oriT* sites near the edges of this peak that point inward from opposite directions. Most of the second, less-defined peak in introgression scores seems to fit moderately well with the other two Hfr donors, REL296 and REL291, whose *oriT* sites are at about 1.3 and 0.8 Mbp, respectively, with the former transferring in the direction of the peak introgression scores and the latter transferring in the same direction toward the broad shoulder between about 0.1 and 1 Mbp. However, the other shoulder of the second, less defined peak—from about 1.2 to 1.5 Mbp—is not explained by the logic of Hfr donor transfer. We also find donor-specific markers in some recombinant clones near the various donor-specific auxotrophic mutations (Fig 2B), but almost all of these nearby introgressions involved a different donor. Nonetheless, the near absence of introgression along the circular chromosome between approximately 1.8 and -1 Mbp—representing over 40% of the genome—fits quite well with the Hfr donor *oriT* sites and directionality. On balance, then, patterns of introgression provide strong, albeit imperfect, support for Hypothesis 1.

We also found compelling evidence of strong purifying selection at the sites of the auxotrophy mutations in the K-12 Hfr donors. Recall that these mutations mean that the cells cannot produce essential amino acids and so cannot grow and persist in the minimal medium used for the STLE. Therefore, purifying selection should remove any recombinant cells that

acquired any of the donors' auxotrophy mutations. The two donors with transfer properties that can account well for the introgression peak between about -1 and 0 Mbp have auxotrophic mutations located at positions that would sharpen the peak by limiting introgression at each edge. In particular, REL288 has an auxotrophy mutation in the *ilvD* gene that lies just beyond the *oriT* site for REL298; and REL298 has an auxotrophy mutation in the *argA* gene that lies a short distance past the *oriT* site for REL288. The other two Hfr donors, REL291 and REL296, have auxotrophy mutations in the *argE* and *leuB* genes, respectively, that would contribute to the observed decline in introgression scores on the broader shoulder of the less defined peak from about 0 to 1.5 Mbp on the circular map. (REL298 also has a second auxotrophy mutation in *leuB*, but this gene is very far from its *oriT* site and thus probably not relevant to the observed patterns of introgression.) On balance, we also find support for Hypothesis 3A, whereby selection against the effectively lethal auxotrophy mutations in the donor strains reinforces and sharpens the patterns of introgression generated by the mechanics of gene transfer according to Hypothesis 1.

Hypothesis 3B offers a different selection-based explanation for the patterns of introgression. It rests on the idea that selection should also act against donor alleles in those genes where beneficial mutations arose in the LTEE and were present in a particular recipient at the start of the STLE. If this hypothesis were correct, then we would expect to see few, if any cases, where these beneficial mutations were removed and replaced by donor DNA. The evidence in support of this hypothesis is ambiguous, at best, because of the considerable variation among the recombinant clones, in terms of both the proportion of their DNA that comes from the donor strains and the extent to which the LTEE-derived beneficial mutations have been retained or replaced. Also, replacements of LTEE-derived beneficial mutations might simply reflect the recent introgression of donor alleles into the recipients (i.e., shortly before the STLE ended)—alleles that would eventually go extinct if conjugation were stopped. In STLE population Ara-1, only 1 of the 9 presumed beneficial mutations present in the recipient was replaced by donor DNA (in both recombinant clones), but ~22% of the recombinant genomes was donor DNA (Fig 1). This pattern is consistent with Hypothesis 3B. By contrast, consider population Ara+1: 4 of the 5 presumed beneficial mutations in the recipient were replaced, but only ~32% of the two recombinant genomes came from the donors (Fig 1). Across the 12 populations, there is a slight tendency for these presumed beneficial mutations to have been replaced by donor alleles more often than the average genomic site, contrary to this hypothesis. On balance, the evidence does not support Hypothesis 3B.

We can also exclude a hypothetical scenario in which beneficial LTEE-derived alleles in the recipients were generally replaced by K-12 alleles that were as or more beneficial. We examined the recombinant clone sequences to determine whether these replacements reverted the gene to its pre-LTEE ancestral state (i.e., the corresponding sequence in REL606, including the case in which that sequence is identical in REL606 and K-12) or, alternatively, introduced a different allele. As usual, we summarize the results for the odd-numbered final clones, but the results do not differ substantively for the even-numbered clones. We examined alignments of 60 proteins (from both non-mutator and hypermutator clones) containing LTEE-evolved alleles in the recipients that were replaced by recombination with the donors. Of those proteins, 11 changed to the K-12 donor state that differs from the pre-LTEE REL606 state; 37 changed to the K-12 donor state that is identical to the pre-LTEE REL606 state; 8 changed back to the pre-LTEE REL606 state that differs from the K-12 donor state; and 4 changed to new states comprising combinations of K-12, REL606, and new mutations.

Fig 3 illustrates the four cases in which new alleles emerged. The *yghJ* gene in population Ara-1 was evidently affected by at least three recombination events: the whole gene is derived from K-12 except for two regions, (spanning residues 803–824 and 1068–1212 in the alignment)

genes where mutations repeatedly reached high frequency in populations with the ancestral mutation rate were under positive selection (Table 1). We therefore closely examined the alignments of the 30 proteins containing LTEE-evolved mutations in the recipients that were replaced by recombination in the nine odd-numbered non-mutator clones. Recombination with the K-12 donors changed 5 LTEE-evolved alleles into the donor state that differs from the pre-LTEE state; changed 18 LTEE-evolved alleles into the donor state that is identical to the pre-LTEE state (i.e., the same sequence is present in REL606 and K-12); changed 4 LTEE-evolved alleles back to the pre-LTEE state that differs from the donor state; and changed 3 LTEE-derived alleles into new alleles (Table 2). Eleven of the 22 cases where an LTEE-evolved allele went back to its pre-LTEE ancestral state occurred in genes under strong positive selection in the LTEE [21, 27], indicating that many beneficial mutations were removed by

Table 2. Loci containing nonsynonymous mutations in odd-numbered recipient clones from non-mutator populations that were replaced by donor alleles.

Population	Gene name	Allelic state of gene after HGT	Strong positive selection in LTEE
Ara+1	<i>arcA</i>	K-12	Yes
Ara+1	<i>hslU</i>	New allele	Yes
Ara+1	<i>ycfC</i>	K-12 (REL606)	No
Ara+1	<i>yecE</i>	K-12 (REL606)	No
Ara+4	<i>infB</i>	K-12 (REL606)	Yes
Ara+4	<i>mreB</i>	K-12 (REL606)	Yes
Ara+4	<i>spoT</i>	K-12	Yes
Ara+5	<i>rplF</i>	K-12 (REL606)	Yes
Ara+5	<i>rpsC</i>	K-12 (REL606)	No
Ara+5	<i>yhdG</i>	K-12 (REL606)	Yes
Ara-1	<i>malT</i>	K-12	Yes
Ara-1	<i>rpsM</i>	K-12 (REL606)	No
Ara-1	<i>yghJ</i>	New allele	No
Ara-3	<i>fis</i>	K-12 (REL606)	No
Ara-3	<i>hslU</i>	K-12 (REL606)	Yes
Ara-3	<i>malT</i>	K-12	Yes
Ara-3	<i>mrdB</i>	K-12 (REL606)	No
Ara-3	<i>pykF</i>	K-12 (REL606)	Yes
Ara-3	<i>rpsG</i>	K-12	No
Ara-3	<i>tdcR</i>	K-12 (REL606)	No
Ara-3	<i>topA</i>	K-12 (REL606)	Yes
Ara-4	<i>garP</i>	K-12 (REL606)	No
Ara-4	<i>infA</i>	K-12 (REL606)	No
Ara-4	<i>mreB</i>	K-12 (REL606)	Yes
Ara-4	<i>nadR</i>	REL606	Yes
Ara-4	<i>pykF</i>	New allele	Yes
Ara-4	<i>spoT</i>	REL606	Yes
Ara-4	<i>ydfI</i>	REL606	No
Ara-4	<i>ypdC</i>	REL606	No
Ara-5	<i>ybaL</i>	K-12 (REL606)	Yes

Note: Allelic state indicates whether the sequence in the recombinant clone is identical to the K-12 but not the REL606 sequence (“K-12”); identical to both the K-12 and the REL606 sequences (“K-12 (REL606)”); identical to the REL606 but not the K-12 sequence (“REL606”); or is a patchwork of K-12, LTEE, and new mutations (“New allele”). Donor alleles often only partially replaced recipient alleles; see Fig 3 for examples and S1 Dataset for details.

<https://doi.org/10.1371/journal.pgen.1007199.t002>

recombination with the donors. The new alleles that evolved after recombination had effectively reconstructed the ancestral states in *hslU* in Ara-1 and *pykF* in Ara-4 imply that these reversions of beneficial LTEE-derived mutations by recombination were not adaptive.

Gene conversion and new mutations

The following analyses focus, for simplicity, on the odd-numbered recombinant clones from the STLE populations that were not hypermutable; however, the even-numbered clones are similar. We noticed that these recombinant clones often had more new mutations than did typical LTEE clones that had evolved for 1000 generations [21]. When we looked for evidence of parallel evolution among these new mutations, we found strong but spurious signals in two genes, *nohB* and *waaQ*; in particular, these genes had multiple identical mutations in multiple lineages. The likely explanation for these parallel changes is gene conversion, in which recombination occurred between non-orthologous genes in the K-12 donors and B recipients, such as between diverged paralogs or perhaps even non-homologous sequences. To investigate further the possibility of gene conversion, we scored all genes that had three or more new mutations in the same recombinant genome as putative gene conversion events (Table 3). Because so many apparently multi-mutation events occurred, and usually in multiple lineages, we think each case is best explained by a single gene-conversion event, not by multiple mutations in the same gene.

We also used data on positive selection on specific genes in the LTEE [21] to ask whether new nonsynonymous mutations in the STLE tended to occur in the same genes. Genes affected by nonsynonymous mutations resulting from putative gene conversions had a mean *G*-score of 0, while the genes harboring all other new nonsynonymous mutations had a mean *G*-score of 33.76. This difference, though only marginally significant (two-sided Welch's *t*-test, *p*-value = 0.035), suggests that nonsynonymous mutations that arose during the STLE were under stronger positive selection than those that occurred by non-homologous recombination.

One of our most puzzling findings is that many LTEE-derived mutations, including some that were almost certainly beneficial, were lost in the recombinant clones from the Ara+1 and Ara-4 STLE populations (Fig 1). These genes were among those under strong positive selection for new mutations in the LTEE [21], and the STLE environment was almost the same as the LTEE. The Ara+1 and Ara-4 STLE populations account for 12 of 30 nonsynonymous replacements, and

Table 3. Putative gene-conversion events in recombinant genomes.

Gene annotation	Lineages mutated	Mutation count	Positions mutated
<i>ECB_00510/nohB</i>	3	21	7
<i>essQ</i>	1	6	6
<i>gmhB/rrsH</i>	3	8	3
<i>nohB</i>	3	15	5
<i>rhsB</i>	2	4	2
<i>rhsC</i>	1	3	3
<i>rrlA</i>	1	3	3
<i>thrW/yagP</i>	3	3	1
<i>waaQ</i>	3	16	6
<i>yafM/fhiA</i>	1	12	12
<i>yahG/yahI</i>	2	4	2
<i>ydfQ</i>	1	4	4
<i>ydfU</i>	1	7	7
<i>yjgL</i>	2	15	9

<https://doi.org/10.1371/journal.pgen.1007199.t003>

14 of 16 new nonsynonymous mutations after excluding the multisite gene-conversion events. This association suggests that the loss of beneficial mutations to recombination in these lineages also led to stronger selection for new beneficial mutations elsewhere in the genome.

In contrast, the Ara-3 STLE population had 8 nonsynonymous replacements, but no new nonsynonymous mutations to compensate. The Ara-3 recombinants are the only ones with genomes that derive primarily from the K-12 donor strains (Fig 1, S2 Fig). Also, this population underwent unexpected changes in its ecology, which led to a substantial decline in its fitness relative to the common competitor used in the STLE [17] and the emergence of a negative frequency-dependent interaction between different recombinant genotypes, probably caused by a crossfeeding interaction [19]. These striking changes in genetic background and ecological context may well have altered the genetic targets of selection in this population.

No characteristic distribution of lengths of recombinant segments

We examined the distributions of recombinant segment lengths to see whether conjugation left a consistent signature in this respect. The left column of Fig 4 shows the distribution of lengths of DNA segments derived from the K-12 donor in the recombinant clones, and the right column shows the length distribution of the B-derived segments (using the odd-numbered clones). To explore the possibility of characteristic segment lengths, we excluded Ara-3 (which was almost entirely donor-derived), Ara-6 and Ara+2 (which had little or no donor DNA), and the three mutator lineages (Ara-2, Ara+3, Ara+6) because DNA repair processes also affect recombination. Even using only the remaining six lineages, the recombinants show significant heterogeneity in the length distribution of their donor-derived segments (Kruskal-Wallis rank sum test, chi-squared = 27.297, df = 5, $p < 10^{-4}$). As noted before, K-12 specific markers occur densely over the REL606 reference genome (S1 Fig), and so the distribution of markers cannot account for this heterogeneity. It is unclear what accounts for this variability across the recombinant lineages. It might reflect differences in the initial recipient genotypes, including their receptivity to conjugation and recombination, or early events in the history of the STLE populations that affected their subsequent receptivity to these processes.

Rate of genomic change due to recombination versus mutation

Table 4 compares the number of synonymous changes introduced by recombination to those caused by mutation. These data exclude probably spurious inferences of synonymous changes introduced by gene conversion events (Table 3). We divided new synonymous changes based on whether they were found in recipient- or donor-derived segments. Even if we exclude the synonymous changes in the donor-derived segments, the rate at which the non-hypermutable recombinant lineages accumulated synonymous changes over the 1000 generations of the STLE was several times higher than the corresponding rate during the LTEE [21, 28–30]. This higher rate presumably reflects some mutagenic effect of recombination, such as error-prone repair of double-strand breaks during integration of donor DNA into a recombinant genome. The fact that there are almost as many new synonymous mutations in the donor-derived as in the recipient-derived segments, despite the smaller cumulative target size of the former regions, also supports that interpretation. In any case, the vast majority of synonymous changes that arose during the STLE were introduced through recombination, not by mutation. Twenty of the 24 sequenced recombinant clones acquired hundreds, thousands, or even tens of thousands of synonymous changes by recombination, whereas no clone, even in the mutator lineages, had as many as 20 synonymous mutations (Table 4). The four exceptional recombinant clones are those that, as described earlier, acquired little or no donor DNA (including both clones from STLE population Ara+2 and single clones from Ara+3 and Ara-6).

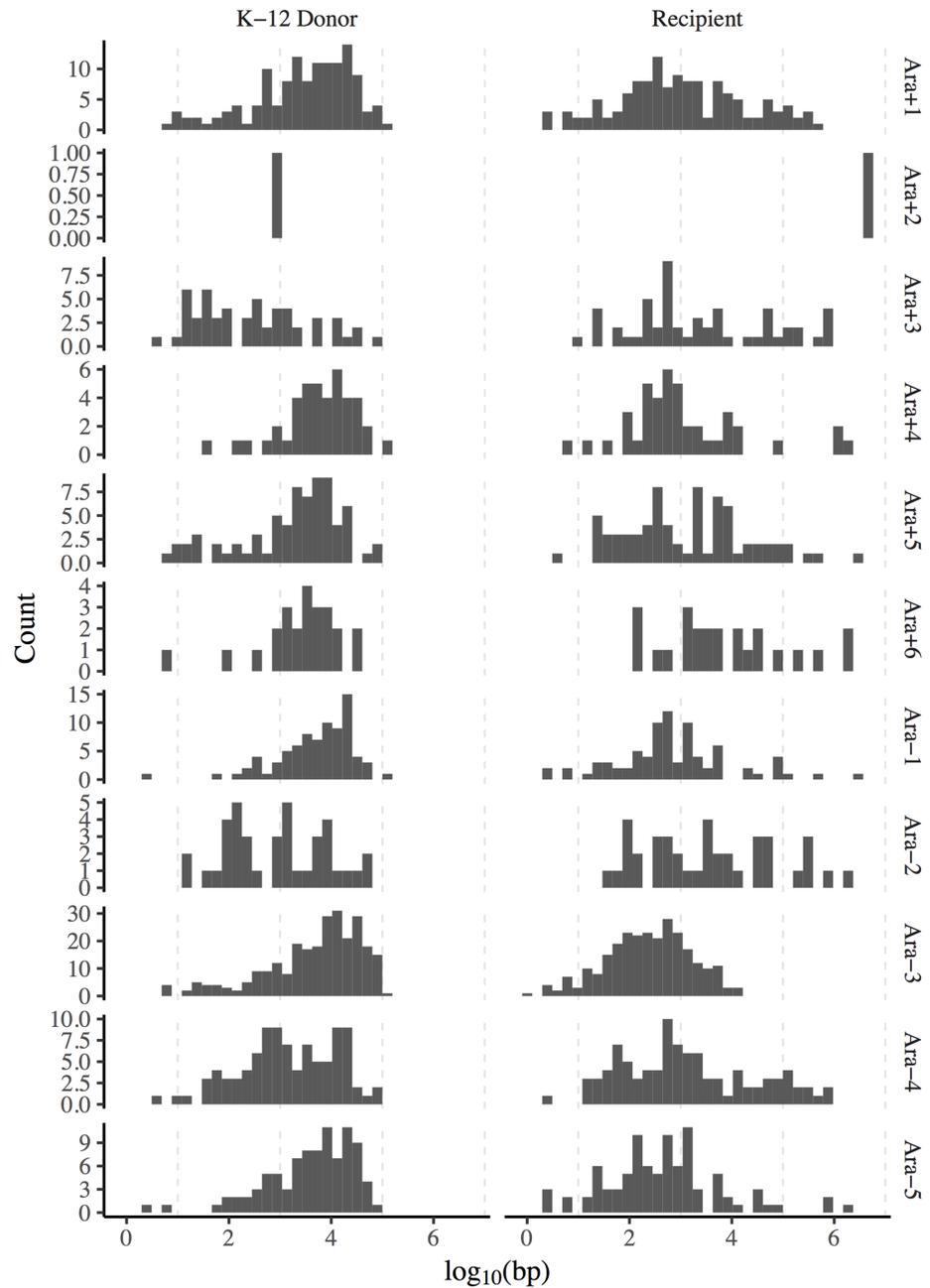


Fig 4. The length distributions, shown on a logarithmic scale, of DNA segments derived from donors (left column) and recipients (right column) in the odd-numbered recombinant genomes.

<https://doi.org/10.1371/journal.pgen.1007199.g004>

Excluding those four atypical clones and others from hypermutable populations (Ara+3, Ara+6, and Ara-2), the ratio of synonymous changes introduced by recombination and mutation is generally well over 1000 (Table 4). For the three hypermutable populations, the ratio of synonymous changes introduced by recombination and mutation is still high at ~70–90. A recent estimate of the ratio of changes observed in recombinant regions relative to spontaneous mutations in natural *E. coli* populations is ~10 [31], and so the intergenomic recombination rate in the STLE was clearly much greater than the recombination rate in nature.

Table 4. Synonymous changes, including replacements from donors and new mutations, in recombinant genomes after 1000 generations of the STLE.

Population (Clone)	Recombined synonymous replacements	Synonymous mutations in recipient-derived segments	Synonymous mutations in donor-derived segments	Recombined / mutated synonymous changes
Ara+1 (REL11734)	8788	2	2	2197–4394
Ara+1 (REL11735)	8766	2	2	2191.5–4383
Ara+2 (REL11736)	0	0	0	ND
Ara+2 (REL11737)	0	1	0	0.0
Ara+3* (REL11738)	1	13	0	0.08
Ara+3* (REL11739)	1369	18	1	72.1–76.1
Ara+4 (REL11740)	5228	0	1	5228 –ND
Ara+4 (REL11741)	3135	0	0	ND
Ara+5 (REL11742)	3648	0	0	ND
Ara+5 (REL11743)	3670	0	0	ND
Ara+6* (REL11744)	659	9	0	73.2
Ara+6* (REL11745)	656	7	0	93.7
Ara–1 (REL11746)	5839	3	1	1459.7–1946.3
Ara–1 (REL11747)	5825	0	1	5825 –ND
Ara–2* (REL11748)	1142	13	0	87.8
Ara–2* (REL11749)	1141	10	3	87.8–114.1
Ara–3 (REL11750)	23396	3	2	4679.2–7798.7
Ara–3 (REL11751)	23454	0	2	11727 –ND
Ara–4 (REL11752)	3749	0	0	ND
Ara–4 (REL11753)	3730	1	0	3730
Ara–5 (REL11754)	7852	2	2	1963–3926
Ara–5 (REL11755)	6415	1	0	6415
Ara–6 (REL11756)	3197	2	0	1598.5
Ara–6 (REL11757)	0	0	0	ND

Note: Ratios of recombined to mutated synonymous changes are calculated with (lower bound) and without (upper bound) synonymous mutations in donor-derived segments. Only one value is shown when there were no synonymous mutations in the donor-derived segments. ND indicates a ratio is not defined because its denominator equals zero. Hypermutator clones are marked with an asterisk.

<https://doi.org/10.1371/journal.pgen.1007199.t004>

The conjugative F plasmid persisted in population Ara–3

One unexpected finding is the predominantly donor-derived ancestry of the recombinant clones from STLE population Ara–3 (Fig 1, S2 Fig and S3 Fig). We considered the possibility

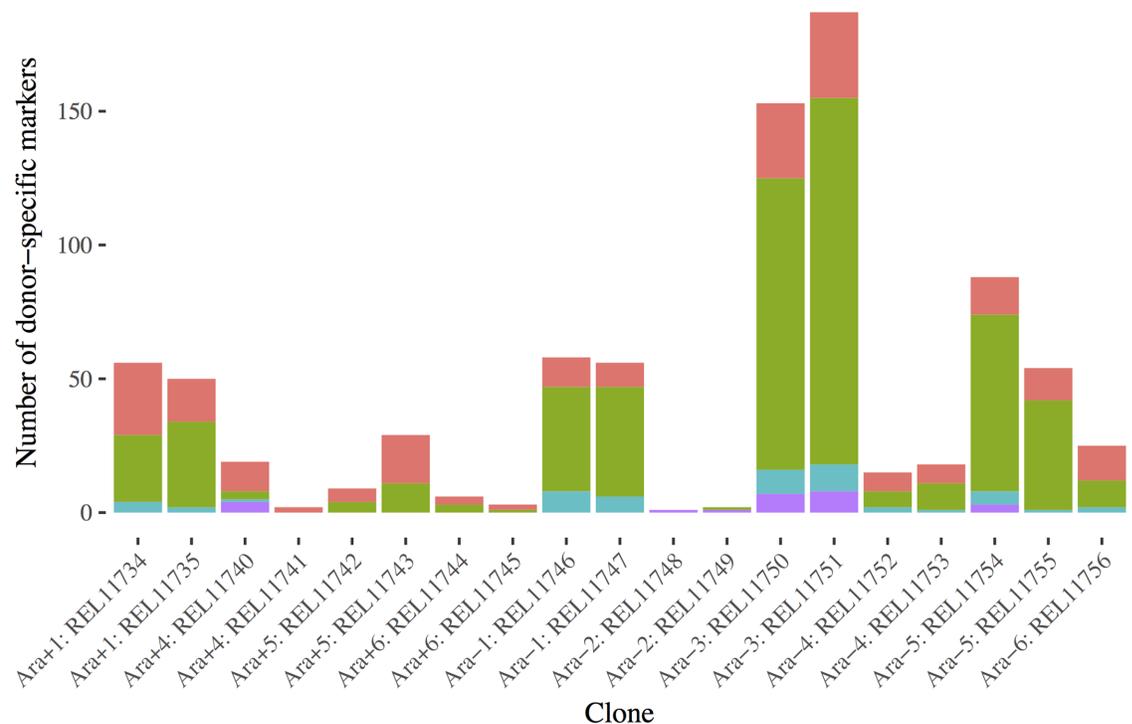


Fig 5. The number and provenance of donor-specific genetic markers in recombinant clones. REL288 markers are shown in red, REL291 markers in green, REL296 markers in teal, and REL298 markers in purple. Four clones have very little or no donor DNA, and another has no donor-specific markers.

<https://doi.org/10.1371/journal.pgen.1007199.g005>

that this ancestry might reflect the reversion of an auxotrophy mutation in one of the K-12 donors, which might have allowed it to persist and perhaps rise to dominance. If that were the case, then we would expect its ancestry to stem from just one of the four donor strains. Instead, the recombinant clones from Ara-3 contain genetic markers from all four donors as well as the B-derived recipient. In fact, all of the STLE populations (except Ara+2) contain markers derived from at least two donors (Fig 5). In any case, the Ara-3 population does not show evidence of being descended from a single donor strain that somehow survived and displaced the recipient population.

If the F plasmid persisted in this population, either in a donor-derived lineage or by the conversion of a recipient into a secondary donor, then conjugation might have occurred above and beyond the recombination treatment imposed every fifth day during the STLE. Although no donor-derived lineage became established, it is possible that a recipient was converted to an F plasmid-carrying donor, despite the shaking that occurred after 1 hour of the recombination treatment, which was expected to interrupt conjugation [17]. Indeed, more recent studies have found that F plasmid-mediated conjugation readily occurs in *E. coli* at the 120 rpm shaking speed of the STLE, albeit under different conditions [32, 33]. We found sequencing reads that map to the entire F plasmid in both Ara-3 clones, but not in any other clones we sequenced (S4 Fig). We do not know whether these Ara-3 clones contained free copies of the plasmid or, alternatively, had integrated the plasmid into the chromosome, which would make them potential Hfr donors. The F-plasmid contains repetitive sequences and insertion elements with homology to multiple locations on the *E. coli* chromosome, making integration a possibility. Souza *et al.* (1997) reported that the Ara-3 clones had become resistant to tetracycline, like the Hfr donors and unlike the recipients or any of the other recombinants they tested.

However, we saw no reads that mapped to the TetR gene, and we do not know the reason for this discrepancy.

STLE continuation experiment

One potentially confounding factor in our analysis is that some donor segments might have been introduced during one of the final rounds of the conjugation treatment of the STLE, and hence they could be deleterious variants that are present only transiently and destined for extinction. To test that possibility, we restarted the 12 STLE populations from their final samples and propagated them for an additional 30 days (200 generations) without further addition of the donor strains. This period would thus allow time for recently generated maladapted recombinants to be outcompeted by more-fit members of the respective populations. We then sequenced whole-population samples from the starting and final time points (S2 Table). If the introgressed K-12 alleles were deleterious, then we would expect their frequency to decrease over time. By measuring how the frequency of K-12 donor alleles changed after stopping conjugation, we assessed whether these alleles were usually evolutionary dead-ends or, instead, persisted by positive selection on introgressed alleles or linked beneficial mutations.

Donor-derived alleles increased in frequency in several populations during the STLE continuation (Fig 6), although some populations, most notably Ara+5, showed a mixture of directional changes. Therefore, the introgressed donor-derived alleles do not appear to impose a consistent fitness burden on the recombinant populations. However, we cannot rule out the possibility that some transferred alleles impose a fitness cost, but nevertheless hitchhiked with beneficial mutations elsewhere in the recombinant genomes [34]. Also, many donor-derived alleles might have increased in frequency as neutral hitchhikers; if a given allele was neutral, then the probability that it would hitchhike to fixation (assuming the STLE continuation was of sufficient duration) should be equal to its initial frequency. Some K-12 donor alleles might even have conferred a selective advantage; the existence of at least a few such alleles is suggested by the tendency for clusters of points to lie above the diagonal. Even though the continuation experiment shows that donor alleles on the whole did not impose a large fitness burden on the recombinant populations, it does not exclude the possibility that recurrent bouts of recombination during the STLE helped drive some maladaptive alleles to fixation.

The continuation experiment and associated metagenomic sequencing also allowed us to infer haplotypes that had fixed before the end of the STLE (Fig 7). Specifically, we identified derived alleles with frequency equal to 1 at the start and finish of the continuation experiment (corresponding to generations 1000 and 1200, respectively, with respect to the STLE) that were also present in both recombinant clones isolated from a given population. We saw that reversions of beneficial LTEE-evolved mutations went to fixation in several of the STLE populations. In the three hypermutator populations (Ara+3, Ara+6, and Ara-2), the introgression of donor alleles might have simultaneously removed some deleterious mutations that arose during the LTEE [29]. Nonetheless, it is puzzling to see the recombination-mediated reversion of some clearly beneficial mutations even in non-mutator populations. In some cases (*hslU* in Ara+1, and *pykF* in Ara-4), new mutations became established, presumably after the replacement of the beneficial LTEE-evolved mutations. In other cases (*nadR* in Ara+1 and *spoT* in Ara+4 and Ara-4), beneficial mutations reverted to their pre-LTEE state, and these reversions went to fixation. In general, the locations of parallel introgression events in the inferred LCA of the STLE closely follow the results based on the sequenced clones (S6 Fig).

Given that the F plasmid was found in the Ara-3 STLE clones, we also looked for it in the whole-population metagenome samples taken at the start and end of the continuation experiment. The F plasmid was present in the final samples from the continuation experiment in two

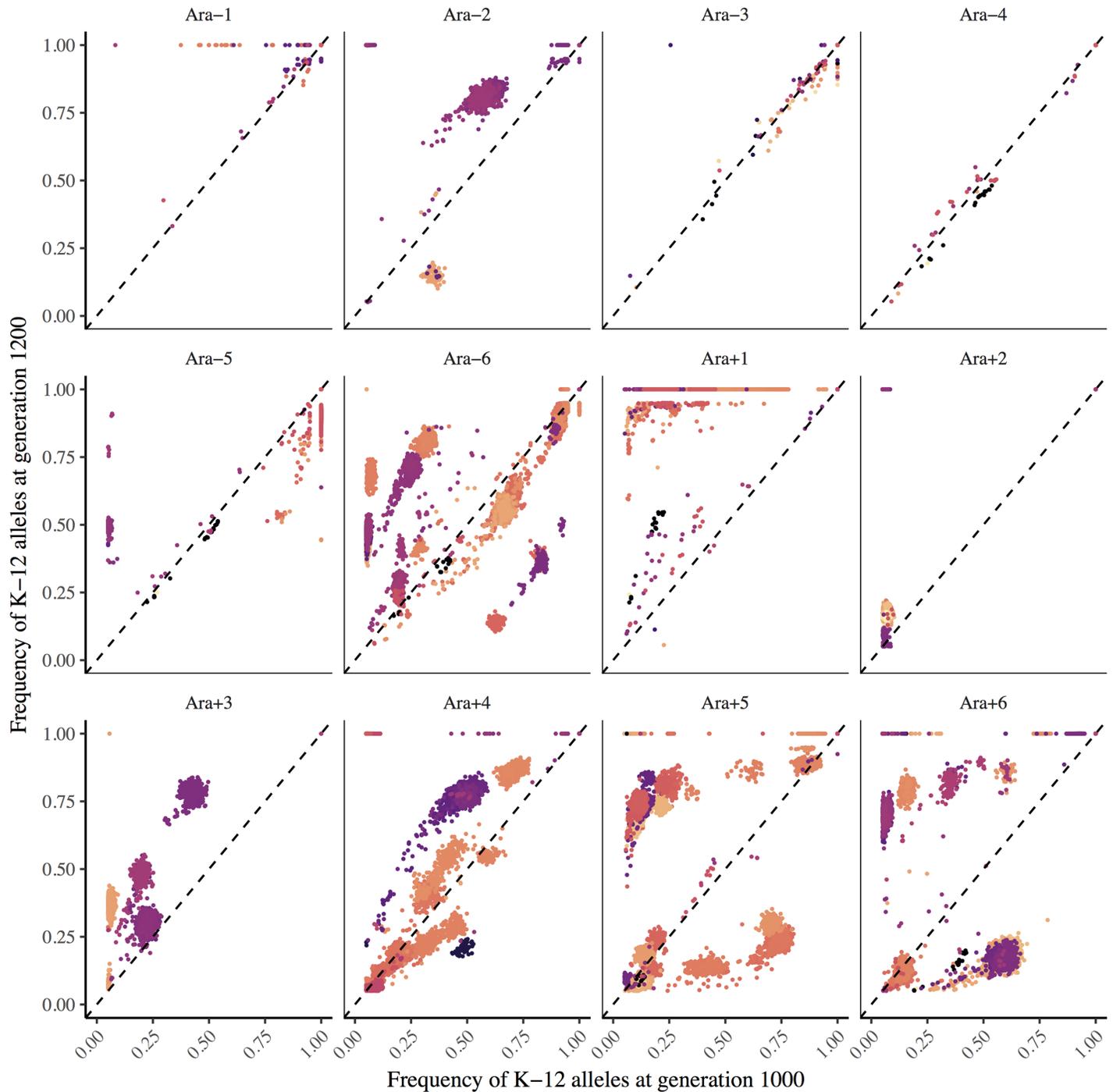


Fig 6. Alleles derived from the K-12 donors tended to increase in frequency during the STLE continuation experiment. Initial and final K-12 allele frequencies are plotted on the x- and y-axis, respectively. Alleles that increased in frequency lie above the dashed diagonal line, and those that decreased lie below the diagonal. Alleles are colored based on their genomic position, so clusters with the same color probably belong to the same haplotype.

<https://doi.org/10.1371/journal.pgen.1007199.g006>

populations, Ara+1 as well as Ara-3 (S7 Fig). To get a sense of the frequency of the plasmid in these two populations, and how it changed during the continuation experiment, we compared sequence read coverage of the plasmid and on the chromosome in the initial and final samples.

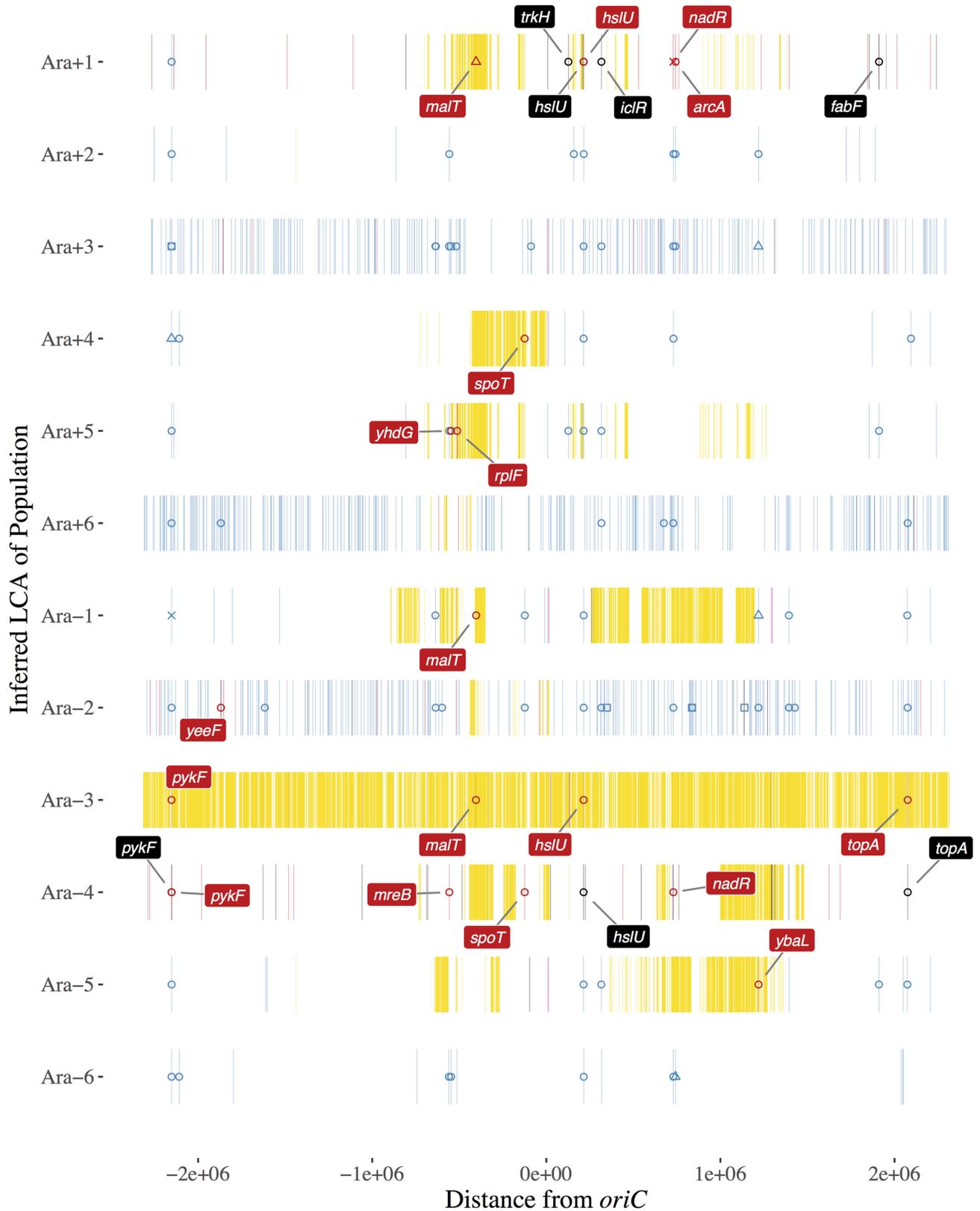


Fig 7. Mutations at 100% frequency in both the initial (1000 generation) and final (1200 generation) samples of the STLE continuation experiment and in both recombinant clones, which we infer to have been present in the last common ancestor (LCA) of each population. See legend to Fig 1 for description of symbols and labels.

<https://doi.org/10.1371/journal.pgen.1007199.g007>

Table 5. Estimated frequency of F plasmids in populations Ara+1 and Ara-3 of the STLE.

Population	Generation	Average coverage of chromosome	Average coverage of plasmid	Estimated plasmid frequency
Ara+1	1000	647.8	3	0.0046
Ara+1	1200	686.9	250.3	0.36
Ara-3	1000	730.2	545.1	0.75
Ara-3	1200	433.1	213.6	0.49

Note: Estimated plasmid frequency assumes one chromosome per cell and one plasmid per plasmid-bearing cell. Average coverage is the mean number of sequencing reads that map to the chromosome or plasmid per site.

<https://doi.org/10.1371/journal.pgen.1007199.t005>

In population Ara-3, the estimated plasmid frequency fell from 75% to 49% during the continuation experiment, whereas in Ara+1 it increased from <1% to 36% (Table 5). Even excluding these two populations, donor-derived alleles tended to increase in frequency in several of the other continuation populations (Fig 6), indicating that plasmid persistence was not necessary for that outcome. Overall, these results indicate that introgressed K-12 alleles did not generally impose a sufficient fitness cost to drive down their frequency, even in the absence of further conjugation.

Recurrent conjugation over a sufficiently long period means that some neutral, and perhaps deleterious, donor alleles will almost certainly fix in an adaptively evolving recipient population through a ratchet-like hitchhiking process. Suppose that a neutral donor allele is introduced into a genetic background that contains a beneficial mutation undergoing a selective sweep; depending on when the donor allele was introduced relative to the sweeping mutation, the donor allele will reach some frequency p . Then, the probability that the next beneficial sweep drives that neutral donor-derived allele to fixation is p . With probability $1 - p$ the next beneficial mutation occurs on a different genetic background, thereby driving the donor allele to extinction. Nonetheless, the process repeats again and again with each successful transfer of a neutral allele. We suggest that this ratchet-like process might explain some of the unexpected dynamics observed during the STLE.

Discussion

In this study, we sequenced and analyzed donor, recipient, and recombinant genomes from the Souza-Turner-Lenski experiment (STLE). Two main results are clear. First, we found parallel evolution in the genomic structure of recombinant *E. coli*. This pattern appears to be largely explained by the biases caused by the molecular biology of conjugation, coupled with selection against the effectively lethal auxotrophy mutations carried by the donor strains. Second, there was substantial introgression of donor DNA into most of the recipient populations. Indeed, the rate of recombination was sufficiently high that many beneficial mutations, which had previously evolved during 7,000 generations of asexual evolution in the same environment, were erased by recombination. We estimate that the effective recombination rate, expressed relative to the rate of genomic change by new mutations, was at least 100-fold higher in the STLE than previously estimated in nature for *E. coli* [31]. However, if the majority of synonymous mutations in natural isolates of *E. coli* arose during periods of transient hypermutability, then the appropriate comparison would be to the hypermutator populations of the STLE. In that case, the effective recombination rate in the STLE would be ~10-fold higher than the rate for nature.

In addition, and to our surprise, we found that one STLE population, designated Ara-3, had predominantly K-12 donor ancestry. The Ara-3 recombinant clones lack all of the LTEE-derived mutations present in their ancestral recipient, but they still have small segments that

derive from the *E. coli* B progenitor used to start the LTEE. The effect of recombination was so strong in this population that highly beneficial alleles in the recipient clone at the *pykF*, *malT*, *hslU*, and *topA* loci were erased (Fig 7). Moreover, the entire F plasmid was present in both of the sequenced Ara-3 recombinant clones, whereas it was absent, as expected, from the sequenced recombinant clones from all other populations. (The F plasmid was found at a much lower frequency at the end of the STLE in one other population.) The F plasmid clearly became established in the Ara-3 population, but it is unclear whether it did so as a free plasmid or was integrated into a recipient chromosome. Several explanations for the F plasmid's persistence are possible, and they are not mutually exclusive. First, experiments have shown that F⁺ strains can convert F⁻ cells into F⁺ cells during a quasi-epidemic of plasmid transmission [35]. If this occurred in Ara-3, then newly infected F⁺ recipients could transmit the plasmid, along with any donor genes that might hitchhike, to further recipients. Second, a B recipient might have been converted into an Hfr donor and delivered small B segments to a K-12 donor strain that then survived. Third, recombination between the K-12 donor and B recipient genomes might have activated an otherwise latent prophage, leading to virus-mediated transduction in the opposite direction to conjugation. Fourth, a K-12 donor strain might have reverted its auxotrophy mutation, allowing it to grow and persist in the minimal medium of the STLE. It is known that the Tn10-transposon mutagenesis used to construct the donor strains [36] yields unstable genotypes, in which the transposons can excise or move to other locations in the genome. Fifth, one K-12 donor might have recombined with a second K-12 strain (which had perhaps lost its F plasmid and thereby become a recipient) in such a way as to repair the nutritional defect. Sixth, some mutation or mutations in the Ara-3 recipient genome may have allowed for vastly more efficient conjugation and DNA incorporation. However, the two recipient strains with defects in their mismatch repair (Ara+3 and Ara-2) did not have more K-12 ancestry than the other recipients, even though previous research has shown that *E. coli* strains with defective mismatch repair have relaxed homology requirements for molecular recombination [7].

Even excluding Ara-3, we observed a great deal of heterogeneity in the amount of introgressed DNA across the STLE populations and in the lengths of donor tracts (Fig 4). Researchers studying natural transformation in other bacterial species have reported that donor segments often cluster into complex mosaic patterns, perhaps generated by long stretches of DNA being disrupted after their uptake or as the result of heteroduplex segregation and correction [37–40]. Our results accord with these previous reports of fine-scale mosaicism of donor- and recipient-derived regions in recombinant genomes. The lengths of many recombinant segments in the STLE are also consistent with the pervasive transfer of genome fragments ranging from ~40 to ~115 kbp reported in natural populations of *E. coli* [31]. Both generalized transduction and conjugation can produce such long tracts of introgressed DNA, although the relative contribution of these mechanisms to horizontal gene transfer in nature is unknown. Experiments have shown that the spatial separation of donor and recipient strains during growth on surfaces can suppress conjugation [41], whereas conjugation can more readily spread genetic material in the well-mixed liquid environment of the STLE.

Evolution experiments with both bacteria and yeast have shown that intergenomic recombination can sometimes speed up the process of adaptation by natural selection [12, 13]. In contrast, the STLE shows that recombination can sometimes act in a manner more analogous to an extremely elevated mutation rate, leading to neutral and even maladaptive changes. Such an effect is not without precedent; for example, plant populations that have evolved resistance to heavy metals found in patchily distributed mine tailings have also evolved selfing to avoid the genetic load of pollen from nearby metal-sensitive populations [42]. The high density of the introduced donors relative to recipients, the high effective rate of recombination, and the

fact that the recipients but not the donors had adapted to the STLE environment appear to have created a similar situation, in which non-adapted donor genes “rained down” on the locally adapted, LTEE-derived recipients. The most striking evidence that recombination could have maladaptive effects in the STLE was the finding that many beneficial mutations were effectively “erased” by replacement with donor alleles that were the same as the LTEE ancestral state, especially in populations Ara+1, Ara-3, and Ara-4. If most donor alleles were neutral or maladaptive in the environment of the STLE, then it is not surprising that the recombination treatment did not speed up adaptation [17]. What is surprising, though, is the extent to which those alleles could evidently invade and replace better-adapted recipient alleles. These results support the first hypothesis proposed by Souza *et al.* (1997), which postulated that recurrent bouts of conjugation were sufficient to drive recombinant genotypes to high frequency.

On the other hand, the fact that most STLE populations did not decline in fitness, despite having some beneficial mutations erased, leaves open the possibility that some other donor-derived segments harbored beneficial alleles that offset the removal of LTEE-derived beneficial mutations. Also, by allowing the STLE populations to evolve for an additional 200 generations without the conjugation treatment, we showed that haplotypes containing donor-derived segments often increased in frequency (Fig 6), contradicting the hypothesis that they could only persist by ongoing recombination. This result supports the second hypothesis proposed by Souza *et al.* (1997), which postulated that donor alleles hitchhiked to high frequency along with beneficial alleles. However, we observed too many linked donor alleles to distinguish between highly beneficial ‘driver’ mutations and their hitchhiking ‘passengers’. This dynamic is common in large asexual populations, where cohorts of mutations track closely in frequency and go to fixation or extinction together [43–45]. In any case, it appears that donor genes often replaced homologous genes in the recipient populations as a consequence, at least in part, of the transmission advantage produced by horizontal gene transfer.

Materials and methods

Overview of the STLE

The experiments performed by Souza *et al.* (1997) are described fully in that paper. In brief, 12 recombinant populations and 12 control populations were started from clones isolated after 7000 generations of the LTEE (S1 Table). These populations were propagated daily for 1000 generations (150 days) following the same transfer regime and using the same DM25 medium and temperature as the LTEE [46]. However, on day 0 and every fifth day thereafter (~33 generations), a mixture of four K-12 Hfr donor strains (REL288, REL291, REL296, and REL298) was added to the recombination treatment populations and allowed to conjugate for 1 h without shaking. The ratio of donor to recipient cells during the treatment was ~4:1. All four donors were auxotrophic for one or more essential amino acids (arginine, leucine, or isoleucine-valine), so they could transfer their genetic material but not grow and persist in the population.

Isolation of clones for genomic analysis

Samples of all 12 recombination-treatment populations from generation 1000 of the STLE (S1 Table) were revived as follows: 100 µL from each frozen stock were pipetted into 10 mL of LB medium, grown for 24 h, diluted and grown in DM25 for two more 24 h cycles, and then spread on LB agar plates. The two colonies that grew closest to randomly placed marks were re-streaked on LB agar plates and then grown in LB medium. These 24 STLE-derived recombinant clones were stored at -80°C (S1 Table). We sometimes refer to these clones as “odd” and

“even” (based on the freezer identification numbers for each pair) when presenting data for just one clone from each population. We estimate that the growth in LB and two cycles in DM25 added roughly 25 generations. Given the very low mutation rate in nonmutator cells [28] and the small fraction of mutations that have any discernible effect on fitness [29], there is little opportunity for meaningful de novo evolution during the isolation of clones. Owing to these procedures, including the unavoidable freezing and thawing of samples, it is also possible that we did not sample clones representative of the diversity present in the STLE populations at generation 1000. However, that concern falls away because our conclusions do not change if we restrict our analysis of the clones to those mutations that fixed in their respective populations by 1000 generations (Fig 7).

Genome sequencing and analysis

The 4 Hfr donor strains (REL288, REL291, REL296, and REL298), 12 LTEE-derived recipient clones used as ancestors in the STLE, and 24 STLE-derived recombinant clones (two per population) were thawed and grown in LB medium, and samples of genomic DNA were isolated from each one. The genomic DNA was then sequenced on an Illumina MiSeq at the MSU RTSF Genomics Core Facility. We used the breseq (version 0.31) pipeline [47] to analyze the genomes, with the ancestral LTEE strain REL606 as the reference genome unless otherwise specified. We used the gdttools utility in breseq to compute a table that lists the union of all the mutations found in the K-12 donor genomes in comparison to REL606, and to identify mutations specific to each of the donor strains.

A custom python program called label_mutations.py was written that performs several tasks. First, it labels all of the “mutations” (i.e., all genetic differences including those that result from recombination) found in the STLE-derived recombinant genomes relative to the REL606 genome by looking at the same site in the donor, recipient, and recombinant genomes. Nine distinct labels are used, as follows. (i) Mutations that are present in both a recombinant clone and its parent recipient clone (but not in REL606 or the union of donor strains) are labeled as “LTEE mutations”. (ii) Mutations found in both a recombinant clone and the union of K-12 donors are labeled as “K-12 mutations”; they represent horizontally transferred alleles. (iii) Any mutations that are present in a recombinant clone, but not found in either the donors or recipient clone, are labeled as “new mutations”; these are mutations that occurred during the STLE, and which do not appear to involve a donor. (iv) Mutations present in the union of K-12 donors that are not found in the recipient clone are labeled “REL606 mutations;” more precisely, these sites are genetic markers that distinguish the *E. coli* B-derived REL606 strain used to start the LTEE from K-12. S1 Fig shows the distribution of these markers along the *E. coli* B chromosome. (v). Mutations found in a recipient, but not in its derived recombinant, are labeled as “deleted mutations”; these mutations were removed by recombination with the donors or otherwise lost during the STLE. (vi–ix). Mutations specific to each of the four donor strains that are found in the recombinants are labeled as “donor-specific mutations”; these mutations are included in the set of “K-12 mutations” in most of our analyses. The program ignores all sites that are identical between the recombinant, recipient, and donors because they provide no useful information. Our analysis also ignores genomic “islands” that are present in K-12 but not REL606. As a second task, the label_mutations.py program produces a table of the genetic markers that distinguish K-12 and REL606 (see label iv above). Third, this program generates a table of the LTEE-specific mutations found in the recipient genomes (see labels i and v above).

An R script called dissertation-analysis.R makes figures and performs statistical tests using the tables of labeled mutations that label_mutations.py produces.

Manual annotation of specific donor genome features

Each donor strain also has two special features: the mutations that make it an amino-acid auxotroph, and the Hfr transfer origin and orientation (i.e., direction of transfer). The breseq analysis found all the auxotrophy mutations present in the donor strains REL288, REL291, REL296, and REL298. These mutations were generated and annotated by Wanner *et al.* (1986), and the strains were confirmed to be auxotrophs by Souza *et al.* (1997). We checked the genomic locations of the auxotrophy mutations returned by breseq against the original strain annotations and a fine-grained traditional linkage map for *E. coli* K-12 [48] from the Coli Genetic Stock Center (<https://cgsc2.biology.yale.edu/Workingmap.php>, last accessed 7/13/17). We note that the *leuB* auxotrophy mutation in REL298 is a S286L (TCG→TTG) mutation in an NAD-binding region of the protein (<http://www.uniprot.org/uniprot/P30125>, last accessed 7/31/17), rather than an amber nonsense mutation as originally annotated.

We did not resolve the position and orientations of the Hfr transfer origins, because the F plasmid contains repetitive sequences and insertion elements that map to multiple locations on the *E. coli* chromosome. The genomes of K-12 and REL606 are largely syntenic [26], and so we used the K-12 linkage map to place the Hfr *oriT* transfer origin sites in the K-12 donors with respect to their homologous genes in REL606. This mapping of K-12 elements to the REL606 genome is only approximate, but it appears to perform reasonably well. Our annotations using breseq and the K-12 linkage map along with cross-references to annotations from the Coli Genetic Stock Center are in a file called donor_hand_annotations.csv.

Because the F plasmid contains repeats and insertion elements that also map to the *E. coli* chromosome, as noted above, we could not resolve whether the recombinant clones and population samples that contained the F element had it integrated in the chromosome or, alternatively, it existed as a free plasmid.

Calculation of lengths of donor and recipient segments in recombinant genomes

Recombination breakpoints occur somewhere in the interval between donor-specific and recipient-specific markers. A minimum estimate of the length of a donor segment would place the recombination breakpoints at the donor markers on each end, while a maximum estimate would place the breakpoints at the flanking recipient markers. In fact, the true breakpoints cannot be known exactly. Our approach uses the minimal estimate on the left, but the maximal estimate on the right, and so it will produce intermediate values that should tend to an overall average length similar to what would be obtained by averaging the minimum and maximum segment lengths. In particular, our algorithm alternates between K-12 and B-derived segments along the genome, switching states when it reaches the alternate marker type. This algorithm is described in detail in [S1 Text](#).

STLE continuation experiment

We performed the following experiment to determine the fate of introgressed K-12 alleles in the absence of the recombination treatment. We revived the 12 populations of the STLE recombination treatment from the samples that were frozen at 1000 generations. We then propagated them daily for 200 generations (30 days) following the same transfer regime and using the same DM25 medium and other conditions as the LTEE [46]. Every 33 generations (5 days) we froze glycerol stocks of all 12 populations. Finally, we isolated genomic DNA from the initial and final population samples, and sequenced the whole-population metagenomes to see how the genetic variation changed during the 200 additional generations without

recombination. When isolating genomic DNA from these population samples, we took 125 μ L of each glycerol stock, then washed and centrifuged the cells twice in DM0 (Davis Minimal medium without glucose) to remove residual glycerol. We then added 100 μ L of the washed and resuspended cells into a flask containing 9.9 mL of DM100 (the same medium used in the LTEE, except with a glucose concentration of 100 μ g/mL to yield more cells), and we plated the remaining 25 μ L onto a tetrazolium arabinose agar plate to make sure that the washing steps did not cause any unexpected population bottleneck.

Supporting information

S1 Text. Calculation of lengths of donor and recipient segments in recombinant genomes.

(PDF)

S1 Table. The identifying numbers and relationships of the 12 sequenced recipient clones used to start the 12 STLE populations and the 24 recombinant clones isolated at the end of the STLE.

(XLSX)

S2 Table. Population samples used in the STLE continuation and subsequent sequencing.

(XLSX)

S1 Fig. Histogram of the density of K-12 specific markers plotted along the genome coordinates of REL606, the ancestral strain of the LTEE. The numbers were binned over 556 DNA segments that are each 8327 bp in length.

(PDF)

S2 Fig. Genome structure of even-numbered clones from recombinant populations after 1000 generations of the STLE. The REL606 genomic coordinates are shown on the x-axis, centered on the *oriC* origin of replication, and the source populations are shown on the y-axis. Genetic markers are shown as vertical lines, with the color indicating the origin of each marker. Markers specific to K-12 donors are yellow; markers specific to recipient clones are blue; markers in deleted regions are light purple; new mutations that arose during the STLE are black; and LTEE-derived mutations that were replaced by donor DNA during the STLE are red. In addition, symbols indicate mutations in genes under positive selection in the LTEE (Table 1). Open circles indicate nonsynonymous point mutations; open squares are synonymous mutations; open triangles are indels; and x-marks are IS-element insertions. Replaced and new mutations in the genes in Table 1 are labeled by their gene names.

(PDF)

S3 Fig. REL4397 and REL4398 are recombinant clones from STLE population Ara-3 that were previously studied by Turner *et al.* (1996). That paper found a cross-feeding interaction between these clones, which puzzlingly had declined in fitness relative to their progenitor during the STLE. The REL606 genomic coordinates are shown on the x-axis, centered on the *oriC* origin of replication, and the source populations are shown on the y-axis. See legend to S2 Fig for description of symbols and labels.

(PDF)

S4 Fig. Recombinant clones from STLE population Ara-3 only contain the F plasmid.

Read coverage over the plasmid reference sequence is plotted for each STLE recombinant clone. The spikes in read coverage near the start and near 20 Kbp of the plasmid reference are false-positive reads that map onto repeat sequences in the plasmid. (A) All Ara-3 STLE clones

are shown with REL4397 in very light orange, REL4398 in light orange, REL11750 in medium orange, and REL11751 in dark orange. (B) For comparison, the four Hfr donors are shown with REL288 in red, REL291 green, REL296 teal, and REL298 purple. (C) All other STLE recombinant clones are shown in shades of grey.
(PDF)

S5 Fig. Sequence divergence between donors and recipients does not appreciably affect introgression efficiency, but it does affect the location of recombination breakpoints. The recipients are all recently derived from REL606, itself a derivative of *E. coli* B. The REL606 genome is 4,629,812 bp, which we divided into 556 bins that are each 8327 bp long. Each point represents one bin. The x-coordinate for each point shows the number of mutational differences between K-12 and REL606 in the bin. (A) The y-coordinate shows the median number of parallel introgression events (as shown in Fig 2) within that bin, which does not correlate with divergence (Spearman-rank correlation $r = 0.0347$, $P = 0.4183$). (B) There is no difference in divergence between bins with and without parallel introgression events (Kruskal-Wallis rank-sum test, $P = 0.9235$) (C) The y-coordinate shows the number of recombination breakpoints within each bin, which correlates with divergence (Spearman-rank correlation $r = 0.1706$, $P < 0.0001$). (D) Bins with breakpoints have diverged more than bins without breakpoints (Kruskal-Wallis rank-sum test, $P = 0.0199$).
(PDF)

S6 Fig. Parallel fixations of K-12 genetic markers during the STLE. K-12 markers found in both recombinant clones and at 100% frequency in both the initial (1000 generation) and final (1200 generation) samples of the STLE continuation experiment were summed over each population (omitting the Ara-3 population which is almost completely derived from K-12 donor DNA). The locations of auxotroph mutations in the donor genomes are shown as dashed vertical lines, and the location and orientation of the Hfr *oriT* transfer origin sites are labeled below the x-axis. The four donors are colored as in Fig 2.
(PDF)

S7 Fig. The Ara+1 population at generation 1200 of the STLE continuation contains the F plasmid, as does the Ara-3 population. The Ara+1 and Ara-3 populations are shown in red, and all other populations in black. This plot shows the coverage distribution over the plasmid reference sequence for the initial (generation 1000) and final (generation 1200) samples from the continuation populations. The three blank panels are for samples where breseq (version 0.31) indicated insufficient coverage across the F plasmid.
(PDF)

S1 Dataset. Whole-protein alignments of proteins encoded by genes in which mutations present in the recipient were replaced by donor DNA in the odd-numbered STLE recombinant clones.
(TXT)

Acknowledgments

We thank Neerja Hajela for technical assistance, Valeria Souza and Paul Turner for their prior work on the STLE, and Jeff Barrick for advice on breseq and the design of this project. We thank Rosemary Redfield, Joshua Mell, Chris Marx, Valeria Souza, and Paul Turner for valuable discussions at the 2015 Microbial Population Biology Gordon Research Conference. R.M. thanks Debora Marks and Chris Sander for support while writing this paper.

Author Contributions

Conceptualization: Rohan Maddamsetti, Richard E. Lenski.

Data curation: Rohan Maddamsetti.

Formal analysis: Rohan Maddamsetti, Richard E. Lenski.

Funding acquisition: Rohan Maddamsetti, Richard E. Lenski.

Investigation: Rohan Maddamsetti.

Methodology: Rohan Maddamsetti.

Software: Rohan Maddamsetti.

Supervision: Richard E. Lenski.

Visualization: Rohan Maddamsetti.

Writing – original draft: Rohan Maddamsetti, Richard E. Lenski.

Writing – review & editing: Rohan Maddamsetti, Richard E. Lenski.

References

1. Rosen MJ, Davison M, Bhaya D, Fisher DS. Fine-scale diversity and extensive recombination in a quasi-sexual bacterial population occupying a broad niche. *Science*. 2015; 348(6238):1019–23. <https://doi.org/10.1126/science.aaa4456> PMID: 26023139
2. Sarkar SF, Guttman DS. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied and Environmental Microbiology*. 2004; 70(4):1999–2012. <https://doi.org/10.1128/AEM.70.4.1999-2012.2004> PMID: 15066790
3. Moran NA, McLaughlin HJ, Sorek R. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*. 2009; 323(5912):379–82. <https://doi.org/10.1126/science.1167140> PMID: 19150844
4. Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *The ISME journal*. 2016; 10(7):1589–601. <https://doi.org/10.1038/ismej.2015.241> PMID: 26744812
5. Milkman R, Bridges MM. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics*. 1990; 126(3):505–17. PMID: 1979037
6. Levin BR, Lenski RE. Coevolution in bacteria and their viruses and plasmids. In: Futuyma D, Slatkin M, editors. *Coevolution*. Sunderland, MA: Sinauer Associates; 1983. p. 99–127.
7. Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature*. 1989; 342(6248):396–401. <https://doi.org/10.1038/342396a0> PMID: 2555716
8. Tenaille O, Taddei F, Radman M, Matic I. Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Research in microbiology*. 2001; 152(1):11–6. PMID: 11281320
9. Harrison E, Brockhurst MA. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in microbiology*. 2012; 20(6):262–7. <https://doi.org/10.1016/j.tim.2012.04.003> PMID: 22564249
10. Keightley PD, Otto SP. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature*. 2006; 443(7107):89–92. <https://doi.org/10.1038/nature05049> PMID: 16957730
11. Dixit PD, Pang TY, Maslov S. Recombination-driven genome evolution and stability of bacterial species. *Genetics*. 2017; 207(1):281–95. <https://doi.org/10.1534/genetics.117.300061> PMID: 28751420
12. Cooper TF. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS biology*. 2007; 5(9):e225. <https://doi.org/10.1371/journal.pbio.0050225> PMID: 17713986
13. McDonald MJ, Rice DP, Desai MM. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature*. 2016; 531(7593):233–6. <https://doi.org/10.1038/nature17143> PMID: 26909573
14. Hickey DA, Rose MR. The role of gene transfer in the evolution of eukaryotic sex. *The evolution of sex*. 1988:161–75.

15. Curtiss R III. Bacterial conjugation. *Annual Reviews in Microbiology*. 1969; 23(1):69–136.
16. Ma NJ, Moonan DW, Isaacs FJ. Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. *nature protocols*. 2014; 9(10):2285–300. <https://doi.org/10.1038/nprot.2014.081> PMID: 25188631
17. Souza V, Turner P, Lenski R. Long-term experimental evolution in *Escherichia coli*. V. Effects of recombination with immigrant genotypes on the rate of bacterial evolution. *Journal of Evolutionary Biology*. 1997; 10(5):743–69.
18. Fisher RA. *The genetical theory of natural selection: a complete variorum edition*: Oxford University Press; 1999.
19. Turner PE, Souza V, Lenski RE. Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. *Ecology*. 1996; 77(7):2119–29.
20. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*. 1997; 387(6634):703. <https://doi.org/10.1038/42701> PMID: 9192894
21. Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, et al. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*. 2016; 536(7615):165–70. <https://doi.org/10.1038/nature18959> PMID: 27479321
22. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*. 2009; 461(7268):1243–7. <https://doi.org/10.1038/nature08480> PMID: 19838166
23. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*. 2011; 332(6034):1193–6. <https://doi.org/10.1126/science.1203801> PMID: 21636772
24. Vulić M, Lenski RE, Radman M. Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proceedings of the National Academy of Sciences*. 1999; 96(13):7348–51.
25. Daegelen P, Studier FW, Lenski RE, Cure S, Kim JF. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21 (DE3). *Journal of molecular biology*. 2009; 394(4):634–43. <https://doi.org/10.1016/j.jmb.2009.09.022> PMID: 19765591
26. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3) and comparison of the *E. coli* B and K-12 genomes. *Journal of molecular biology*. 2009; 394(4):653–80. <https://doi.org/10.1016/j.jmb.2009.09.021> PMID: 19765592
27. Maddamsetti R, Hatcher PJ, Green AG, Williams BL, Marks DS, Lenski RE. Core Genes Evolve Rapidly in the Long-term Evolution Experiment with *Escherichia coli*. *Genome biology and evolution*. 2017; 9(4):1072–83.
28. Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Médigue C, et al. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3: Genes, Genomes, Genetics*. 2011; 1(3):183–6.
29. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, et al. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proceedings of the National Academy of Sciences*. 2013; 110(1):222–7.
30. Maddamsetti R, Hatcher PJ, Cruveiller S, Médigue C, Barrick JE, Lenski RE. Synonymous genetic variation in natural isolates of *Escherichia coli* does not predict where synonymous substitutions occur in a long-term experiment. *Molecular biology and evolution*. 2015; 32(11):2897–904. <https://doi.org/10.1093/molbev/msv161> PMID: 26199375
31. Dixit PD, Pang TY, Studier FW, Maslov S. Recombinant transfer in the basic genome of *Escherichia coli*. *Proceedings of the National Academy of Sciences*. 2015; 112(29):9070–5.
32. Zhong X, Kro JE, Top EM, Krone SM. Accounting for mating pair formation in plasmid population dynamics. *Journal of theoretical biology*. 2010; 262(4):711–9. <https://doi.org/10.1016/j.jtbi.2009.10.013> PMID: 19835890
33. Wan Z, Varshavsky J, Teegala S, McLawrence J, Goddard NL. Measuring the rate of conjugal plasmid transfer in a bacterial population using quantitative PCR. *Biophysical journal*. 2011; 101(1):237–44. <https://doi.org/10.1016/j.bpj.2011.04.054> PMID: 21723834
34. Good BH, Desai MM. Deleterious passengers in adapting populations. *Genetics*. 2014; 198(3):1183–208. <https://doi.org/10.1534/genetics.114.170233> PMID: 25194161
35. Lin A, Jimenez J, Derr J, Vera P, Manapat ML, Esvelt KM, et al. Inhibition of bacterial conjugation by phage M13 and its protein g3p: quantitative analysis and model. *PLoS one*. 2011; 6(5):e19991. <https://doi.org/10.1371/journal.pone.0019991> PMID: 21637841
36. Wanner BL. Novel regulatory mutants of the phosphate regulon in *Escherichia coli* K-12. *Journal of molecular biology*. 1986; 191(1):39–58. PMID: 3540312

37. Lin EA, Zhang X-S, Levine SM, Gill SR, Falush D, Blaser MJ. Natural transformation of *Helicobacter pylori* involves the integration of short DNA fragments interrupted by gaps of variable size. *PLoS pathogens*. 2009; 5(3):e1000337. <https://doi.org/10.1371/journal.ppat.1000337> PMID: 19282979
38. Mell JC, Shumilina S, Hall IM, Redfield RJ. Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS pathogens*. 2011; 7(7):e1002151. <https://doi.org/10.1371/journal.ppat.1002151> PMID: 21829353
39. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. A high-resolution view of genome-wide pneumococcal transformation. *PLoS pathogens*. 2012; 8(6):e1002745. <https://doi.org/10.1371/journal.ppat.1002745> PMID: 22719250
40. Mell JC, Lee JY, Firme M, Sinha S, Redfield RJ. Extensive cotransformation of natural variation into chromosomes of naturally competent *Haemophilus influenzae*. *G3: Genes, Genomes, Genetics*. 2014; 4(4):717–31.
41. Freese PD, Korolev KS, Jiménez JI, Chen IA. Genetic drift suppresses bacterial conjugation in spatially structured populations. *Biophysical journal*. 2014; 106(4):944–54. <https://doi.org/10.1016/j.bpj.2014.01.012> PMID: 24559997
42. Antonovics J. Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity*. 2006; 97(1):33–7. <https://doi.org/10.1038/sj.hdy.6800835> PMID: 16639420
43. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*. 2013; 500(7464):571–4. <https://doi.org/10.1038/nature12344> PMID: 23873039
44. Maddamsetti R, Lenski RE, Barrick JE. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics*. 2015; 200(2):619–31. <https://doi.org/10.1534/genetics.115.176677> PMID: 25911659
45. Buskirk SW, Peace RE, Lang GI. Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proceedings of the National Academy of Sciences*. 2017; 114(31):8330–5. <https://doi.org/10.1073/pnas.1702314114> PMID: 28720700
46. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist*. 1991; 138(6):1315–41.
47. Deatherage DE, Traverse CC, Wolf LN, Barrick JE. Detecting rare structural variation in evolving microbial populations from new sequence junctions using breseq. *Frontiers in genetics*. 2015; 5:468. <https://doi.org/10.3389/fgene.2014.00468> PMID: 25653667
48. Berlyn MK. Linkage map of *Escherichia coli* K-10: the traditional map. *Microbiology and Molecular Biology Reviews*. 1998; 62(3):814–984. PMID: 9729611